# HOW TO PROJECT CUSTOMER RETENTION

**PETER S. FADER AND BRUCE G. S. HARDIE**

MARKETPLACE

**PETER S. FADER**
is the Frances and Pei-Yuan Chia Professor of Marketing, The Wharton School, The University of Pennsylvania, Philadelphia, PA; e-mail: faderp@wharton.upenn.edu

**BRUCE G. S. HARDIE**
is Associate Professor of Marketing, London Business School, London, United Kingdom; e-mail: bhardie@london.edu

The authors thank Michael Berry and Gordon Linoff for providing the data used in this article, and Naufel Vilcassim for his helpful comments. The second author acknowledges the support of the London Business School Centre for Marketing and the hospitality of the Department of Marketing at the University of Auckland Business School.

At the heart of any contractual or subscription-oriented business model is the notion of the retention rate. An important managerial task is to take a series of past retention numbers for a given group of customers and project them into the future to make more accurate predictions about customer tenure, lifetime value, and so on. As an alternative to common "curve-fitting" regression models, we develop and demonstrate a probability model with a well-grounded "story" for the churn process. We show that our basic model (known as a "shifted-beta-geometric") can be implemented in a simple Microsoft Excel spreadsheet and provides remarkably accurate forecasts and other useful diagnostics about customer retention. We provide a detailed appendix covering the implementation details and offer additional pointers to other related models.



JOURNAL OF INTERACTIVE MARKETING VOLUME 21 / NUMBER 1 / WINTER 2007
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/dir.20074

## INTRODUCTION

A defining characteristic of a contractual or subscription business setting is that the departure of a customer is observed. For example, the customer has to contact the firm to cancel a mobile phone contract; similarly, a local theater company can observe that a patron has not renewed an annual subscription.[1] As such, it makes sense to talk of metrics such as retention and churn rates: The retention rate for Period $t$ ($r_t$) is defined as the proportion of customers active at the end of Period $t-1$ who are still active at the end of Period $t$, and the churn rate for a given period is defined as the proportion of customers active at the end of Period $t-1$ who dropped out in Period $t$.[2]

As we seek to understand the nature of customer behavior in a contractual setting, it is useful to draw on the survival analysis literature. One particularly useful concept for characterizing the distribution of customer lifetimes is that of the *survivor function*, denoted by $S(t)$, which is the probability that a customer has "survived" to Time $t$ (i.e., is still active at $t$). Recalling the definition of a retention rate, it follows that

$$\begin{aligned} S(t) &= r_1 \times r_2 \times \cdots \times r_t \\ &= \prod_{i=1}^{t} r_i, \end{aligned} \tag{1}$$

which implies

$$r_t = \frac{S(t)}{S(t-1)}. \tag{2}$$

Several quantities of managerial interest can be easily calculated directly from the survivor function. For example, the expected (or average) tenure of a customer is simply the area under the survivor function. In a discrete-time setting, this is computed as

$$\text{expected tenure} = \sum_{t=0}^{\infty} S(t).$$

In light of (1), the standard textbook expression for (expected) customer lifetime value (CLV) in a contractual setting that (correctly) reflects the phenomenon of nonconstant retention rates,

$$E(CLV) = \sum_{t=0}^{\infty} m \left\{ \prod_{i=1}^{t} r_i \right\} \left( \frac{1}{1+d} \right)^t,$$

can be written as

$$E(CLV) = \sum_{t=0}^{\infty} m \frac{S(t)}{(1+d)^t}.$$

In a contractual setting, the empirical survivor function $\widehat{S(t)}$ is simply the proportion of customers acquired at Time 0 who are still active at Time $t$. A major problem in using the empirical survivor function to compute expected tenure or lifetime value is that the observed time horizon is often quite limited. Suppose we observe a particular cohort of customers over their first 5 years with the firm, which implies we can compute $\widehat{S(1)}, \ldots \widehat{S(5)}$ (By definition, $\widehat{S(0)} = 1$.) The quantity $\widehat{S(0)} + \cdots + \widehat{S(5)}$ is the expected customer lifetime for the members of the cohort over this period. Similarly, we can compute expected CLV during the first 5 years of a customer's relationship with the firm; however, we would be underestimating the expected tenure and CLV of a new customer, as we would be ignoring the remaining life of those cutomers who are active at the end of Year 5. To compute the true expected tenure and CLV, we need to be able to project the survivor function beyond the observed time horizon. That is, we need to create estimates of $S(6), S(7), \ldots$ given the data $\widehat{S(1)}, \ldots \widehat{S(5)}$. This projected survivor function also is needed if we wish to compute the expected residual tenure or lifetime value of an individual who has been a customer for, say, 3 years.

An obvious approach is to fit some flexible function of time to the observed data. Then resulting regression equation can be used to project the survivor function beyond the range of observations, from which we

[1] This is in contrast to a noncontractual setting, a defining characteristic of which is that the departure of a customer is not observed by the firm (see "Limits to Application" section for a discussion of the implications of this characteristic).

[2] Strictly speaking, we should talk of retention and churn *probabilities*, not *rates*.

can compute expected tenure, CLV, and so on. In a popular book on data mining, Berry and Linoff (2004) explored this idea (pp. 392–393); their conclusion regarding the viability of such an exercise is evident in the title of their sidebar discussion "Parametric approaches do not work."

The objective of this article is to present an alternative approach to the problem of projecting the survivor function—one that does "work." We formulate a probabilistic model of contract duration that is based on a simple story of customer behavior. The resulting model offers useful diagnostic insights and is very easy to implement using Microsoft Excel.

In the next section, we replicate and extend Berry and Linoff's (2004) analysis. We then present a simple probability model of customer lifetime and demonstrate the value of using a formal model to predict future customer behavior. We conclude with a discussion of several issues that arise from this work.

## PROJECTING SURVIVAL USING SIMPLE FUNCTIONS OF TIME

The survival data presented in Table 1 are for two segments of customers ("Regular" and "High End") for an unspecified subscription-type business. These data were presented in graphical form in Berry and Linoff (2004, chap. 12). The High End data were used by Berry and Linoff in their examination of parametric approaches to the projection of the survivor function.

**TABLE 1** Observed %Customers Surviving at Least 0–12 Years

| | %SURVIVING | |
|---|---|---|
| YEAR | REGULAR | HIGH END |
| 0 | 100.0 | 100.0 |
| 1 | 63.1 | 86.9 |
| 2 | 46.8 | 74.3 |
| 3 | 38.2 | 65.3 |
| 4 | 32.6 | 59.3 |
| 5 | 28.9 | 55.1 |
| 6 | 26.2 | 51.7 |
| 7 | 24.1 | 49.1 |
| 8 | 22.3 | 46.8 |
| 9 | 20.7 | 44.5 |
| 10 | 19.4 | 42.7 |
| 11 | 18.3 | 40.9 |
| 12 | 17.3 | 39.4 |

Suppose we only have the first 7 years of data and wish to compute estimates of $S(8)$, $S(9)$, . . . . If we were to give these data to a student who had just completed a typical data analysis course, the natural starting point would be to fit a linear function of time to the data and use the resulting regression equation to project the survivor function over the future periods. Recognizing that the data are not linear, some students would add a quadratic term to try to capture the curvature in the data. More sophisticated students would specify some nonlinear function of time, such as an exponential function.

In their "Parametric approaches do not work" sidebar, Berry and Linoff (2004) estimated and compared this set of regression models with the following results:[3]

| | | |
|---|---|---|
| Linear | $y = 0.925 - 0.071t$ | $R^2 = 0.922$ |
| Quadratic | $y = 0.997 - 0.142t + 0.010t^2$ | $R^2 = 0.998$ |
| Exponential | $\ln(y) = -0.062 - 0.102t$ | $R^2 = 0.963$ |

where $y$ is the proportion of customers surviving at least $t$ years. These equations then are used to extrapolate the survivor function to Year 12; Figure 1 recreates the plot presented in Berry and Linoff's sidebar (p. 393).

The fit of all three models up to and including Year 7 is reasonable, and the quadratic model provides a particularly good fit. But when we consider the projections beyond the model calibration period, all three models break down dramatically. The linear and exponential models underestimate Year 12 survival by 81 and 30%, respectively, while the quadratic model overestimates Year 12 survival by 92%. Furthermore, the models lack logical consistency: The linear model would have $S(t) < 0$ after year 14, and according to the quadratic model the survivor

[3] In the models run by Berry and Linoff (2004), time is indexed 1, 2, . . . , 8, but to maintain consistency with the definitions of $S(t)$ discussed earlier [specifically $S(0) = 1$], we reindex time to 0, 1, . . . , 7. This has no impact at all on the fit or forecasting performance of any of the models.

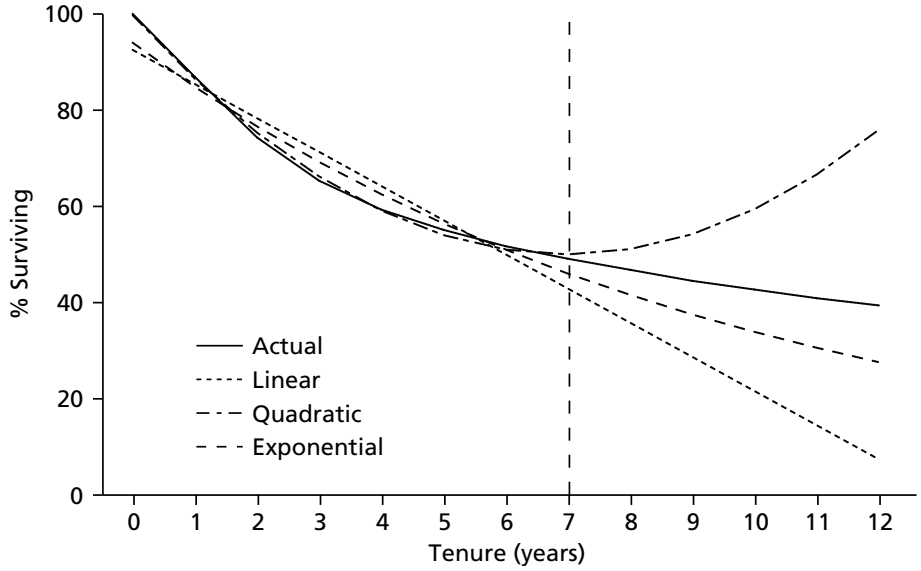


**FIGURE 1**

Actual Versus Regression-Model-Based Estimates of the Percentage of High End Customers Surviving at Least 0–12 Years

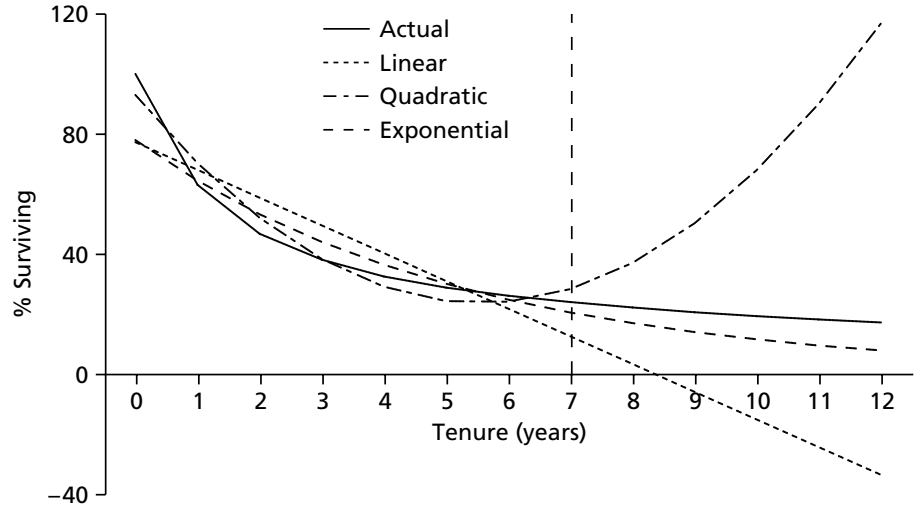


**FIGURE 2**

Actual Versus Regression-Model-Based Estimates of the Percentage of Regular Customers Surviving at Least 0–12 Years

function will start to increase over time, which is not possible. It is therefore not surprising that Berry and Linoff (2004) concluded that parametric curves do not "work" for the task of projecting the survivor function over time.

Repeating this analysis for the Regular segment yields the following equations:

| | | |
|---|---|---|
| Linear | $y = 0.773 - 0.092t$ | $R^2 = 0.776$ |
| Quadratic | $y = 0.930 - 0.249t + 0.022t^2$ | $R^2 = 0.960$ |
| Exponential | $\ln(y) = -0.248 - 0.190t$ | $R^2 = 0.915.$ |

and the corresponding fits and projections are reported in Figure 2. The projections associated with the linear and quadratic models are terrible and illogical once again. The exponential model does not appear to

be very bad in the figure, but in fact it underestimates Year 12 survival by 54%. This is not an acceptable range of error.

Of course, we could try out different arbitrary functions of time, but this would be a pure curve-fitting exercise at its worst. Furthermore, it is hard to imagine that there would be any underlying rationale for the equation(s) that we might settle upon. Faced with this situation, it is tempting to "throw up our hands" in despair and say that we cannot project the survivor function beyond the range of observations.

However, we feel that such a conclusion is premature. After all, in other areas of marketing there are plenty of models that have been used to provide accurate forecasts of the behavior of a cohort of customers beyond the range of observations (e.g., Hardie, Fader, & Wisniewski, 1998, for the case of new-product-sales forecasting). Thus, in the next section, we formulate a probabilistic model of contract duration that is based on a simple "story" of customer behavior.

## A DISCRETE-TIME MODEL FOR CONTRACT DURATION

Consider the following story of customer behavior in a contractual setting:

- At the end of each period, a customer flips a coin: "heads" she cancels her contract, "tails" she renews it.
- For a given individual, the probability of a coin coming up "heads" does not change over time.
- $P$("heads") varies across customers.

Of course, people do not make their contract renewal decisions on the basis of coin flips; rather, this story is a paramorphic representation of customer behavior. The third element of the story should not be controversial, as the notion of heterogeneity is central to marketing; however, some readers might find the second element contrary to their expectation that retention rates increase over time as the customer gains more experience with the product or service. But rather than overcomplicate our story, we start with the simplest possible set of assumptions and only add supposedly richer "touches of reality" if the model does not "work." As seen shortly, no additional assumptions will be required in this particular case.

To operationalize this verbal model, we need to translate the elements of this story into the language of mathematics. More formally, our proposed model for the duration of customer lifetimes is based on the following two assumptions:

1. An individual remains a customer of the firm with constant retention probability $1 - \theta$. This is equivalent to assuming that the duration of the customer's relationship with the firm, denoted by the random variable $T$, is characterized by the (shifted) geometric distribution with probability mass function and survivor function

$$P(T = t \mid \theta) = \theta(1 - \theta)^{t-1}, \quad t = 1, 2, 3, \ldots \tag{3}$$

$$S(t \mid \theta) = (1 - \theta)^t, \qquad t = 1, 2, 3, \ldots. \tag{4}$$

2. Heterogeneity in $\theta$ follows a beta distribution with pdf

$$f(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)}, \ \alpha, \beta > 0,$$

where $B(\cdot\,, \cdot)$ is the beta function.

The assumption of geometrically distributed lifetimes follows from the first two elements of our simple story of customer behavior; it is perfectly consistent with the sequential coin-flip description. The beta distribution will be less familiar to most readers, but it is a very reasonable way to characterize heterogeneity in the churn probabilities because it is a flexible distribution that is bounded between zero and one. If one thinks about how the "coin-flip" probabilities are likely to vary across individuals, there are four principal possibilities, as illustrated in Figure 3. If both parameters of the beta distribution ($\alpha$ and $\beta$) are small ($<1$), then the mix of churn probabilities is "U-shaped," or highly polarized across customers. If both parameters are relatively large ($\alpha, \beta > 1$), then the probabilities are fairly homogeneous. Likewise, the distribution of probabilities can be "J-shaped" or "reverse-J-shaped" if the parameters fall within the remaining ranges as shown in the figure. It is not essential for the reader to remember all of these cases, but these parameters can offer useful diagnostics to help the manager understand the degree (and

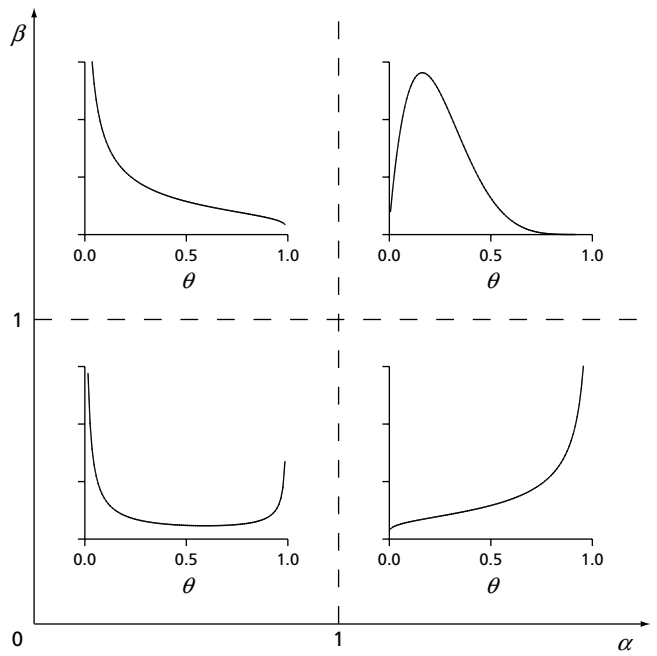


**FIGURE 3**
General Shapes of the β Distribution as a Function of α and β

nature) of heterogeneity in churn probabilities across the customer base.

Given these two model assumptions, how can we compute the probability that a customer fails to renew his contract at the end of Period $t$ or survives beyond Period $t$ [$P(T = t)$ and $S(t)$, respectively]? Since this customer's value of θ is unobserved, we cannot use (3) and (4). We therefore take the expectation of (3) and (4) over the beta distribution that characterizes the cross-sectional heterogeneity in θ to arrive at the corresponding expressions for a randomly chosen individual:

$$P(T = t|\alpha, \beta) = \frac{B(\alpha + 1, \beta + t - 1)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots \quad (5)$$

$$S(t|\alpha, \beta) = \frac{B(\alpha, \beta + t)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots \quad (6)$$

(The mathematically inclined reader is referred to Appendix A for step-by-step details of the derivations.) We call this model the shifted-beta-geometric (sBG) distribution. Nonbusiness applications of this model include the number of menstrual cycles required to achieve pregnancy (Weinberg & Gladen, 1986) and the length of stays in a psychiatric hospital (Kaplan, 1982). Direct-marketing applications of related models are discussed later.

We note that while (5) and (6) are expressed in terms of beta functions, we can implement the model without ever having to deal with beta functions directly. As formally derived in Appendix A, we can compute sBG probabilities by using the following forward-recursion formula from $P(T = 1)$:

$$P(T = t) = \begin{cases} \dfrac{\alpha}{\alpha + \beta} & t = 1 \\ \dfrac{\beta + t - 2}{\alpha + \beta + t - 1} P(T = t - 1) & t = 2, 3, \ldots \end{cases} \quad (7)$$

Recall from (2) that the retention rate is the ratio of sequential values of the survivor function. Substituting (6) into (2) and simplifying (see Appendix A) gives us the following expression for the (aggregate) retention rate associated with sBG model:

$$r_t = \frac{\beta + t - 1}{\alpha + \beta + t - 1} \quad (8)$$

Given (8), we can go back to the expression given in (1) and compute $S(t)$ without having to deal with any beta functions.

We immediately see that under the sBG model, the retention rate is an increasing function of time, even though the underlying (unobserved) individual-level retention probability is constant. According to this model, there are no underlying time dynamics at the level of the individual customer; the observed phenomenon of retention rates increasing over time is simply due to heterogeneity (i.e., the high-churn customers drop out early in the observation period, with the remaining customers having lower churn probabilities). This well-known "ruse of heterogeneity" (Vaupel & Yashin, 1985) is often overlooked by those attempting to make sense of various aggregate patterns of customer behavior.

We fit the sBG model to the first 7 years of the data presented in Table 1. For the High End segment,

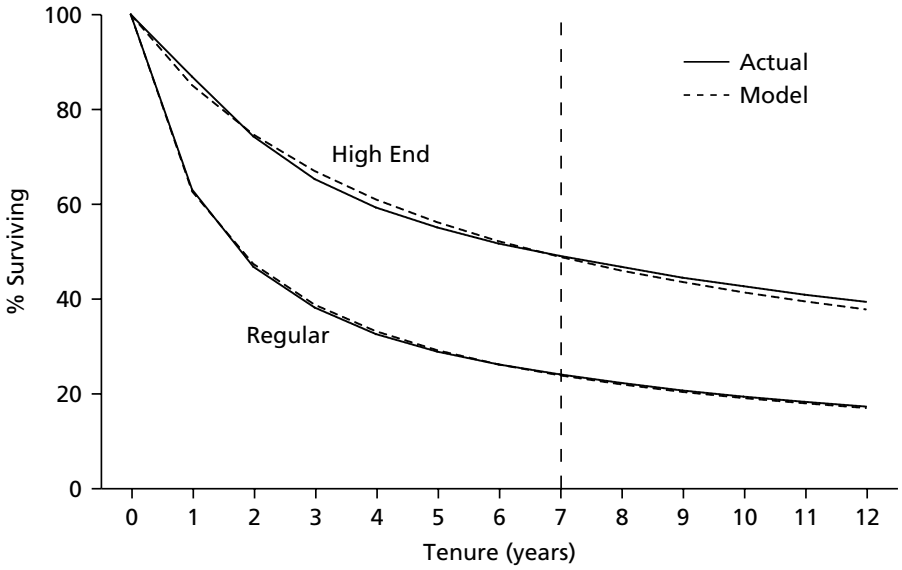


**FIGURE 4**

Actual Versus sBG-Model-Based Estimates of the Percentage of Customers Surviving at Least 0–12 Years for the High End and Regular Segments

$\hat{\alpha} = 0.688$, $\hat{\beta} = 3.806$; for the Regular segment, $\hat{\alpha} = 0.704$, $\hat{\beta} = 1.182$. (See Appendix B for details of how to estimate the model parameters in the familiar Microsoft Excel environment.) Using these parameter estimates, we extrapolate the survivor function for each segment to Year 12. These model-based numbers are plotted in Figure 4, along with the corresponding empirical survivor functions. The resulting predictions are almost too good to be true; the sBG model overestimates Year 12 survival by only 4% and 2% for the High End and Regular segments, respectively. Even though this model is no more complicated than the regression models discussed earlier, its carefully constructed "story" makes it possible to tease out, and therefore accurately project, the critical behavioral components.

Another plot of interest shows the (aggregate) retention rate as a function of tenure. The model-based retention rate numbers [as computed using (8)] are plotted in Figure 5, along with the corresponding observed retention rates as computed from the empirical survivor functions. For both segments, the sBG model accurately tracks the empirical retention rate curves. On one hand, this might not seem surprising since $r_t$ and $S(t)$ are so closely related; on the other hand, however, $r_t$ is harder to predict accurately since it does not have to accumulate across periods as $S(t)$ does, and therefore it is more sensitive to period-to-period variations. Despite the existence of certain unexplained "blips" as in Year 2 for the High End segment, the tracking/prediction plot for $r_t$ is very impressive through Year 12, and there is every reason to believe that the model would continue to perform well over an even longer future horizon.

For both segments, note that the retention rates are an increasing function of the length of a customer's relationship with the firm. The important point to emphasize, once again, is that the sBG "story" assumes that these apparent dynamics are simply a result of heterogeneity; any given individual has a constant (but unknown) retention probability $1 - \theta$. Unlike the conventional wisdom about customer retention, it is *not* a story of individual customers becoming increasingly loyal as they develop a deeper relationship with the firm, and so on. Thus, the observed phenomenon of increasing retention rates is simply a sorting effect in a heterogeneous population (i.e., the high-churn customers drop out early in the observation period, with the remaining customers having lower churn probabilities).

As a final demonstration of the usefulness of the sBG model, we show and contrast the mixing distributions

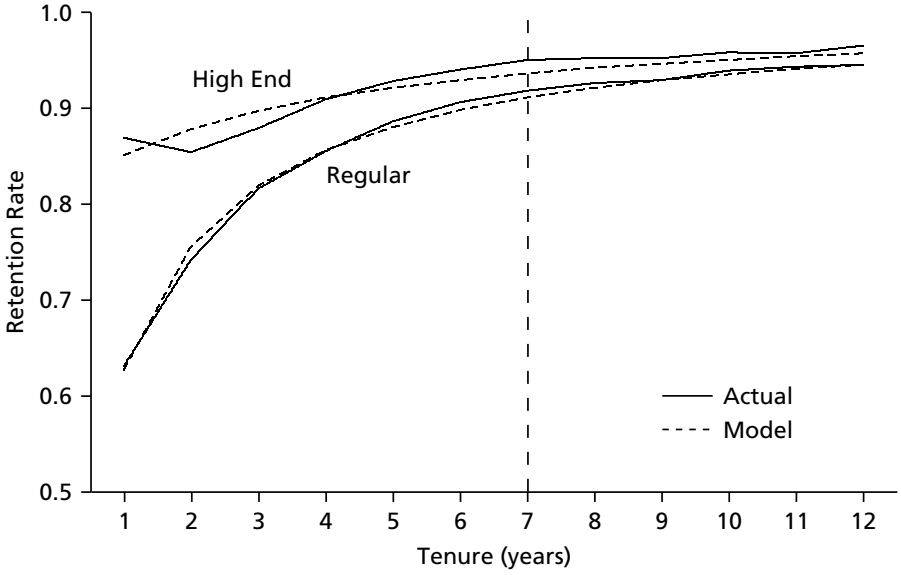


**FIGURE 5**

Actual Versus sBG-Model-Based Estimates of Retention Rates by Tenure for the High End and Regular Segments

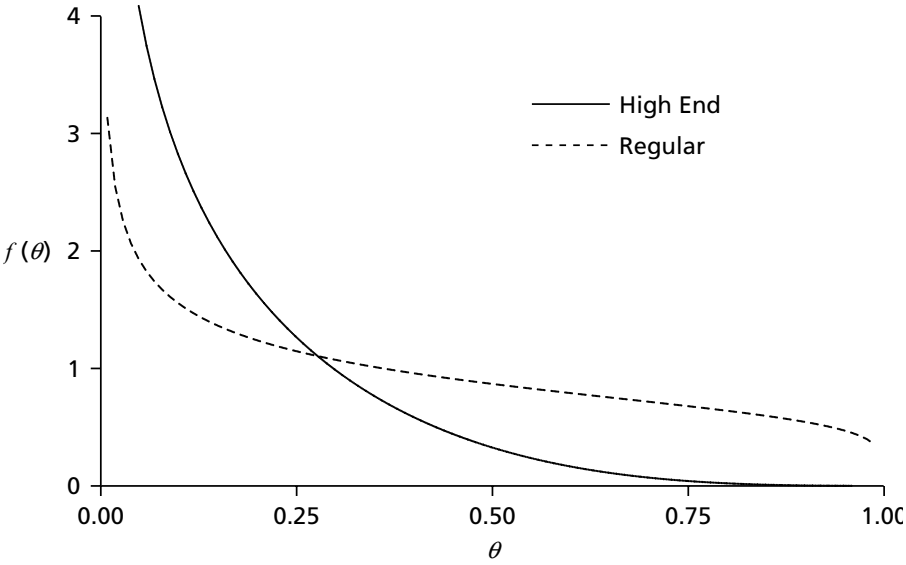


**FIGURE 6**

Estimated Distributions of Churn Probabilities for the High End and Regular Segments

that characterize how the churn probabilities ($\theta$) differ across the individuals in each segment. In Figure 6, we see that both distributions are "reverse-J-shaped." This implies that within each group, most customers have fairly low churn probabilities, but there is a sizeable subsegment within each one that will tend to depart very quickly. These patterns suggest that there is a fairly high degree of heterogeneity within each segment; therefore, a model that does not take these cross-customer differences into account will not perform very well, particularly in terms of out-of-sample forecasting. A closer examination shows

that the overall "weight" of the distribution for the Regular group is shifted slightly to the right compared to that of the High End distribution. This reflects the fact that the Regular group has a higher mean churn probability [$E(\theta) = \alpha/(\alpha + \beta) = 0.37$] compared to that of the High End group [$E(\theta) = 0.15$]. It should be clear from Figures 4 and 5 that this kind of difference in the means exists, but this plot provides a better idea about the nature of these differences at a more fine-grained level.

## DISCUSSION

We have presented the sBG distribution as a model for the duration of customer relationships in a discrete-time contractual setting, and demonstrated that it can provide accurate forecasts and other useful diagnostics about customer retention. Furthermore, we have argued that it is preferable to use such a model instead of arbitrary functions of time. In closing, we discuss limits to its application, related models in the direct-marketing literature, possible extensions to the basic model, and some practical implementation issues.

### *Limits to Application*

The practical problem that drove the development of this model is a desire to project an empirical survivor function (and therefore retention rates) beyond the observed time horizon of our dataset. The ability to perform this projection is central to any attempt to compute CLV or other metrics such as expected tenure if we wish to avoid the "truncation" problem associated with computing these quantities using just the observed survival data. For this particular problem, this simple model should be the first tool the researcher pulls out of his toolkit.

There are other churn-related problems where this should not be the case. In particular, there is a broad literature on churn modeling in which logit models (and far more sophisticated statistical models and data-mining methodologies) are used to determine the correlates of churn (Berry & Linoff, 2004; Parr Rud, 2001). The resulting models then can be used to identify which customers are at risk of churning in the next period so that retention-oriented marketing resources can be targeted at them. Many of the covariates included in these models will vary from period to period (e.g., number of contacts with the customer-service department), and changes in these variables can be strong predictors of customer defection.

However, these models cannot easily be used to address the problem of projecting the survivor function into the future, as we do not have future values of the time-varying covariates. It is therefore important to use the right model for the task at hand, and to acknowledge the limitations to application of any model we develop.

We have referred to the sBG distribution as a model for the duration of customer relationships in a discrete-time contractual setting. Many readers will have glanced over the words "discrete-time" and "contractual" without reflecting on their significance, however, they are very important as we seek to understand when and where it is appropriate to use the model presented in this article.

- By "discrete-time," we mean that transactions can occur only at fixed points in time (e.g., the annual renewal cycles for most professional organizations). This is in contrast to continuous-time, where the transactions can occur at any point in time (e.g., the cancelation of basic utility contracts).
- In a "contractual" setting, the time at which the customer becomes inactive is observed (e.g., when the customer fails to renew a subscription). This is in contrast to a "noncontractual" setting, where the absence of a contract or subscription means that the point in time at which the customer becomes inactive is not observed by the firm (e.g., a catalog retailer). The challenge is how to differentiate between a customer who has ended a "relationship" with the firm versus one who is merely in the midst of a long hiatus between transactions.

This leads to a two-dimensional classification of customer bases: opportunities for transactions (continuous vs. discrete) and type of relationship with customers (noncontractual vs. contractual). The model in this article is for just one of the four possible business contexts.

In continuous-time contractual settings, we should not use the sBG model. Rather, we should use its continuous-time analog, the exponential-gamma (EG)

distribution (also known as the Lomax distribution or the "Pareto distribution of the second kind"). Such a model assumes that the duration of an individual customer's relationship with the firm is characterized by the exponential distribution, and that heterogeneity in "departure rates" is captured by a gamma distribution (Hardie et al., 1998; Morrison & Schmittlein, 1980).

Models for noncontractual settings are more complicated because the time at which a customer becomes inactive, and the likelihood that it has occurred at all, must be inferred from the transaction history. For continuous-time noncontractual settings, we have the Pareto/NBD (Schmittlein, Morrison, & Colombo, 1987) and BG/NBD (Fader, Hardie, & Lee, 2005) models while for discrete-time noncontractual settings, we have the BG/BB model (Fader, Hardie, & Berger, 2004).

## *Related Probability Models and Extensions*

"List falloff" is an important phenomenon in direct marketing. The basic idea is that the response rate from the first mailing to a prospect list is usually higher than that of the second mailing, which in turn is higher than that for the third mailing, and so on. Buchanan and Morrison (1988), hereafter BM, presented a simple probability model of list falloff and showed how the model can be used to determine how many more mailings should be sent to a prospect list given the observed response rates for the first two mailings. Their model is based on assumptions similar to those behind the sBG model: (a) Each person responds to a direct-mail solicitation with constant probability $p$, and (b) $p$ varies across the population according to a beta distribution. While BM base their framework on the beta-binomial model, it could have been derived as an sBG model (e.g., the mailing on which the prospect responds to the offer is characterized by the shifted-geometric distribution). As such, it is possible to identify clear relationships between some of the results in this article [e.g., $r_t$ and $S(t)$] and some quantities of interest in a list-falloff setting.

The BM framework was extended by Rao and Steckel (1995) to incorporate (time-invariant) descriptor variables such as age, income, and sex. This is accomplished using the beta-logistic model (Heckman & Willis, 1977), which extends the beta-binomial model by making the model parameters functions of the descriptor variables. By a similar logic, the effects of time-invariant covariates could be incorporated in the sBG model by making $\alpha$ and $\beta$ functions of the descriptor variables. Incorporating the effects of time-varying covariates (e.g., marketing-mix effects, seasonality) is more complicated. The key is to bring in all of these factors at the right level; that is, at the level of the latent parameter of interest (in this case, $\theta$) instead of just "jamming" different covariate effects into a regression-like model (see Schweidel, Fader, & Bradlow, 2006, for a discussion of how to do this in a continuous-time contractual setting.) However, as noted in the last section, we question the value of such an extension given our modeling objective (i.e., projecting the empirical survivor function beyond the observed time horizon of our dataset).

Both the sBG model and its continuous-time analog (i.e., the EG model) are based on the assumption that the commonly observed phenomenon of increasing retention rates is due entirely to heterogeneity; individual-customer-level retention rates are assumed to be constant. If we wish to allow for the possibility of time dynamics at the level of the individual customer, we can no longer characterize the duration of an individual's relationship with the firm using either the shifted-geometric or exponential distributions, both of which have the "memoryless" property (i.e., the probability of survival to $s + t$, given survival to $t$, is the same as the initial probability of survival to $s$). In a continuous-time setting, we can accommodate this effect by assuming that individual lifetimes can be characterized by the Weibull distribution, which allows for an individual's risk of canceling a contract to increase or decrease as the length of the relationship with the firm increases. In a discrete-time contractual setting, this leads to the beta-discrete-Weibull (BdW) model (Fader & Hardie, 2006), which is a generalization of the sBG model, while in a continuous-time contractual setting, this leads to a generalization of the EG model, the Weibull-gamma (WG) model (Hardie et al., 1998; Morrison & Schmittlein, 1980).

## *Implementation Issues*

Our treatment of how to estimate the sBG model parameters (Appendix B) assumes that we are fitting the model to data for just one cohort of customers. But in practice, we will frequently have data for more

than one cohort, where cohorts are defined by time of acquisition (and possibly acquisition channel, product class, etc.) When faced with data for multiple cohorts, an important model implementation issue is to choose among three possible approaches: (a) to pool the cohorts and estimate a single set of model parameters across them, (b) to estimate a separate set of model parameters for each cohort, or (c) to use a "beta-logistic" version of the sBG with cohort-specific dummy variables. Our decision of how to move ahead is influenced by our beliefs of whether we can view each cohort as the realization of a common underlying contract duration process. The two datasets examined earlier demonstrate that we can expect to see some cross-cohort differences. Schweidel et al. (2006) examined this issue more broadly in a continuous-time setting.

When we have multiple cohorts defined by time of acquisition, the problem with fitting separate models to each cohort is that every new cohort has one less period of information than does its temporal predecessor, which may result in less confidence in the model parameter estimates for the cohorts with fewer data points. The natural starting point in such a situation is to pool the cohorts, assuming that each cohort is the realization of a common underlying contract-duration process, and to estimate one set of parameters using all the data. A more elegant solution would be to add another layer of heterogeneity to the model. That is, we would assume that $\alpha$ and $\beta$ themselves are distributed across cohorts according to some parametric distribution. Using a hierarchical Bayes formulation, this would enable the cohorts with fewer data points to "borrow" information about the possible values of $\alpha$ and $\beta$ from the earlier cohorts rather than relying on the cohort-specific data alone.

## REFERENCES


Berry, M. J. A., & Linoff, G. S. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (2nd ed.). Indianapolis, IN: Wiley.

Buchanan, B., & Morrison, D. G. (1988). A Stochastic Model of List Falloff With Implications for Repeat Mailings. Journal of Direct Marketing, 2(Summer), 7–15.

Fader, P. S., & Hardie, B. G. S. (2006). Customer Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity. Retrieved September 23, 2006, from http://brucehardie.com/papers/022/

Fader, P. S., Hardie, B. G. S., & Berger, P. D. (2004). Customer-Base Analysis With Discrete-Time Transaction Data. Retrieved September 23, 2006, from http://brucehardie.com/papers/020/

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. Marketing Science, 24(Spring), 275–284.

Hardie, B. G. S., Fader, P. S., & Wisniewski, M. (1998). An Empirical Comparison of New Product Trial Forecasting Models. Journal of Forecasting, 17(June–July), 209–229.

Heckman, J. J., & Willis, R. J. (1977). A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women. Journal of Political Economy, 85(February), 27–58.

Kaplan, E. H. (1982). Statistical Models and Mental Health: An Analysis of Records From a Mental Health Center. Unpublished master's thesis, Massachusetts Institute of Technology, Department of Mathematics. Cambridge.

Morrison, D. G., & Schmittlein, D. C. (1980). Jobs, Strikes, and Wars: Probability Models for Duration. Organizational Behavior and Human Performance, 25(April), 224–251.

Parr Rud, O. (2001). Data Mining Cookbook. New York: Wiley.

Rao, V. R., & Steckel, J. H. (1995). Selecting, Evaluating, and Updating Prospects in Direct Marketing. Journal of Direct Marketing, 9(Spring), 20–31.

Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting Your Customers: Who They Are and What Will They Do Next? Management Science, 33(January), 1–24.

Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2006). Modeling Retention in and Across Cohorts. Retrieved September 23, 2006, from http://ssrn.com/abstract=742884

Vaupel, J. W., & Yashin, A. I. (1985). Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. The American Statistician, 39(August), 176–185.

Weinberg, C. R., & Gladen, B. C. (1986). The Beta-Geometric Distribution Applied to Comparative Fecundability Studies. Biometrics, 42(September), 547–560.

# APPENDIX A STEPS IN MODEL DERIVATION

In Appendix A, we walk through the derivations of the key mathematical results presented in this article.

Note the three definitions and results that are central to the derivations that follow.

- The beta function $B(\alpha, \beta)$ is defined by the integral

$$B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta,\ \alpha, \beta > 0. \tag{A1}$$

  Note that $B(\alpha, \beta)$ is simply notation for the definite integral on the right-hand side of (A1).

- The beta function can be expressed in terms of gamma functions:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{A2}$$

- For the purposes of this article, the only thing we need to know about the gamma function is its so-called recursive property:

$$\frac{\Gamma(x+1)}{\Gamma(x)} = x. \tag{A3}$$

***Derivation of (5)***

We derive the sBG expression for $P(T = t)$ in the following manner. If $\theta$ were known, the probability of dropping out in Period $t$ would simply be the shifted-geometric probability $\theta(1-\theta)^{t-1}$. But since $\theta$ is unobserved (and assumed to be distributed randomly across the population), $P(T = t)$ for a randomly chosen individual is the expected value of the shifted-geometric probability of dropping out in Period $t$ (conditional on $\Theta = \theta$), where the expectation is with respect to the beta distribution for $\Theta$, $E[P(T = t \mid \Theta = \theta)]$. (That is, we weight each $P(T = t \mid \Theta = \theta)$ by the probability of that value of $\theta$ occurring, $f(\theta)$.) Since $\Theta$ is a continuous random variable, this is computed as

$$P(T = t \mid \alpha, \beta) = \int_0^1 \underbrace{P(T = t \mid \Theta = \theta)}_{\text{shifted-geometric}} \underbrace{f(\theta \mid \alpha, \beta)}_{\text{beta}}\, d\theta$$

$$= \int_0^1 \theta(1-\theta)^{t-1} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}\, d\theta$$

which, combining terms and moving all non-$\theta$ elements to the left of the integral sign,

$$= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha}(1-\theta)^{\beta+t-2} d\theta.$$

Looking closely at the integral, we see that it is simply the integral expression for the beta function (A1) with parameters $\alpha + 1$ and $\beta + t - 1$. Therefore,

$$P(T = t \mid \alpha, \beta) = \frac{B(\alpha+1, \beta+t-1)}{B(\alpha, \beta)}.$$

[The expression for the sBG survivor function (6) is derived in a similar manner.]

***Derivation of (7)***

To derive the forward-recursion formula used to compute sBG probabilities, first note that

$$P(T = 1 \mid \alpha, \beta) = \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)}$$

which, expressing the beta functions in term of gamma functions (A2),

$$= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \Big/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \Big/ \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta)}.$$

Recalling the recursive nature of the gamma function (A3), $\Gamma(\alpha+1)/\Gamma(\alpha) = \alpha$ and $\Gamma(\alpha+\beta+1)/\Gamma(\alpha+\beta) = \alpha+\beta$. Therefore,

$$P(T = 1 \mid \alpha, \beta) = \frac{\alpha}{\alpha+\beta}.$$

But how does this help us compute $P(T = t)$ for $t = 2, 3, \ldots$? Reflecting on the identity

$$P(T = t) = \frac{P(T = t)}{P(T = t-1)} \times P(T = t-1),$$

if we have a simple expression for the ratio $P(T = t)/P(T = t-1)$, we can easily compute $P(T = 2)$ given the value of $P(T = 1) = \alpha/(\alpha+\beta)$. Given the value of $P(T = 2)$, we can then compute $P(T = 3)$, and so on.

Recalling (5), we have

$$\frac{P(T=t)}{P(T=t-1)} = \frac{B(\alpha+1,\beta+t-1)}{B(\alpha,\beta)} \Big/ \frac{B(\alpha+1,\beta+t-2)}{B(\alpha,\beta)}$$

$$= \frac{B(\alpha+1,\beta+t-1)}{B(\alpha+1,\beta+t-2)}$$

which, expressing the beta functions in term of gamma functions (A2) and canceling terms,

$$= \frac{\Gamma(\beta+t-1)}{\Gamma(\beta+t-2)} \Big/ \frac{\Gamma(\alpha+\beta+t)}{\Gamma(\alpha+\beta+t-1)}$$

which, recalling the recursive nature of the gamma function (A3),

$$= \frac{\beta+t-2}{\alpha+\beta+t-1}.$$

The complete forward-recursion formula naturally follows.

***Derivation of (8)***

We derive the expression for the retention rate as implied by the sBG model by substituting the expression for the sBG survivor function (6) into (2) and simplifying:

$$r_t = \frac{B(\alpha,\beta+t)}{B(\alpha,\beta)} \Big/ \frac{B(\alpha,\beta+t-1)}{B(\alpha,\beta)}$$

$$= \frac{B(\alpha,\beta+t)}{B(\alpha,\beta+t-1)}$$

which, expressing the beta functions in terms of gamma functions (A2) and canceling terms,

$$= \frac{\Gamma(\beta+t)}{\Gamma(\beta+t-1)} \Big/ \frac{\Gamma(\alpha+\beta+t)}{\Gamma(\alpha+\beta+t-1)}$$

which, recalling the recursive nature of the gamma function (A3),

$$= \frac{\beta+t-1}{\alpha+\beta+t-1}.$$

## APPENDIX B IMPLEMENTING THE MODEL IN EXCEL

In Appendix B, we show how to compute the maximum likelihood estimates for the sBG model parameters for the High End dataset using Microsoft Excel. Before providing step-by-step instructions for constructing the worksheet, we briefly review the notion of maximum likelihood estimation.

Suppose we observe a group of $n$ customers for seven periods. Note that $n_1$ customers are "lost" in the first period (i.e., do not renew their contract at the end of that period), $n_2$ in the second period, . . . , with $n_7$ customers not renewing their contracts at the end of the seventh period. It follows that $n - \sum_{t=1}^{7} n_t$ customers are still active at the end of the seventh period.

Assume that the customer lifetimes can be characterized by the sBG distribution. What is the probability that a randomly chosen customer has a lifetime of one period? The answer is the sBG probability $P(T = 1 \mid \alpha, \beta)$. What is the probability that a randomly chosen customer has a lifetime of two periods? The answer is the sBG probability $P(T = 2 \mid \alpha, \beta)$. What is the probability that one randomly chosen customer has a lifetime of one period while another has a lifetime of two periods? Assuming that the propensity of one customer to drop out is independent of the behavior of the other customer, it is simply the product of the respective sBG probabilities: $P(T = 1 \mid \alpha, \beta)P(T = 2 \mid \alpha, \beta)$. It follows that, given specific values of the model parameters $\alpha$ and $\beta$, the joint probability of losing $n_1$ customers in the first period, $n_2$ in the second period, . . ., $n_7$ in the seventh period, and $n - \sum_{t=1}^{7} n_t$ customers still being active at the end of the seventh period is

$$\begin{aligned} P(\text{data} \mid \alpha, \beta) = {} & P(T=1 \mid \alpha, \beta)^{n_1} P(T=2 \mid \alpha, \beta)^{n_2} P(T=3 \mid \alpha, \beta)^{n_3} \\ & \times P(T=4 \mid \alpha, \beta)^{n_4} P(T=5 \mid \alpha, \beta)^{n_5} P(T=6 \mid \alpha, \beta)^{n_6} \\ & \times P(T=7 \mid \alpha, \beta)^{n_7} S(7 \mid \alpha, \beta)^{n - \sum_{t=1}^{7} n_t}. \end{aligned} \quad \text{(B1)}$$

However, we do not know the values of $\alpha$ and $\beta$, even though we believe that the data come from the sBG distribution.

The idea of maximum likelihood estimation is to ask what values of the model parameters maximize the probability (or, more formally, the *likelihood*) of the observed data. We define the likelihood function as

$$\begin{aligned} L(\alpha, \beta \mid \text{data}) = {} & P(T=1 \mid \alpha, \beta)^{n_1} P(T=2 \mid \alpha, \beta)^{n_2} P(T=3 \mid \alpha, \beta)^{n_3} \\ & \times P(T=4 \mid \alpha, \beta)^{n_4} P(T=5 \mid \alpha, \beta)^{n_5} P(T=6 \mid \alpha, \beta)^{n_6} \\ & \times P(T=7 \mid \alpha, \beta)^{n_7} S(7 \mid \alpha, \beta)^{n - \sum_{t=1}^{7} n_t}. \end{aligned} \quad \text{(B2)}$$

and use numerical optimization methods (e.g., the Solver add-in in Excel) to find the values of $\alpha$ and $\beta$ that maximize this function; these are called the *maximum likelihood*

**TABLE B1** Sample Data

| YEAR | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| No. Active | 869 | 743 | 653 | 593 | 551 | 517 | 491 |

*estimates* of the model parameters.[4] As the number computed using (B2) will be very small, we usually work with the natural logarithm of the likelihood function, the so-called log-likelihood function:

$$LL(\alpha, \beta \mid \text{data}) = \ln\left[L(\alpha, \beta \mid \text{data})\right] = \sum_{t=1}^{7} n_t \ln\left[P(T = t \mid \alpha, \beta)\right] + \left(n - \sum_{t=1}^{7} n_t\right) \ln\left[S(7 \mid \alpha, \beta)\right]. \quad \text{(B3)}$$

For a sample of 1,000 High End customers, Table 1 implies the number of customers active at the end of Years 1–7 as reported in Table B1.

Given these data, our task is to "code up" the expression for the model log-likelihood function in an Excel worksheet and find the maximum likelihood estimates of α and β by using Solver to find the values of α and β that maximize the value of this function. The relevant worksheet is shown in Figure B1 and is constructed in the following manner.

- To enter expressions for $P(T = t \mid \alpha, \beta)$without an error message appearing (e.g., `#NUM!` or `#DIV/0!`), we need some "starting values" for α and β. The exact values do not matter—provided they are within the defined bounds (i.e., $\alpha, \beta > 0$)—so we start with 1.0 for α and β, locating these parameter values in Cells `B1:B2`, respectively.
- We enter the values of $t = 1, 2, \ldots, 7$ in Cells `A6:A12`.
- The corresponding values of $P(T = t \mid \alpha, \beta)$ are computed in Cells `B6:B12` using the forward-recursion given in (7):
  - We compute $P(T = 1)$ by entering `=B1/(B1 + B2)` in Cell `B6`.
  - We compute $P(T = 2)$ by entering `=($B$2+A7−2)/($B$1+$B$2+A7−1)*B6` in Cell `B7`.
  - We copy `B7` to `B8:B12`.
- We compute the values of $S(t \mid \alpha, \beta)$ for $t = 1, 2, \ldots, 7$ in Cells `C6:C12`:
  - $S(1)$ is simply $1 - P(T = 1)$, so we enter `=1−B6` in Cell `C6`.
  - For $t > 1$, $S(t) = S(t - 1) - P(T = t)$, so we enter `=C6−B7` in Cell `C7`.
  - We copy `C7` to `C8:C12`.
- The next step is to enter the observed data. The number of customers active at the end of Year 1 ($n = 869$) is entered in cell `D6`, the number for Year 2 ($n = 743$) is entered in cell `D7`, and so on down to 491 customers in cell `D12` for Year 7.
- The number of customers not renewing their contracts each year ($n_t$), as required for the log-likelihood function, is computed in Cells `E6:E12`:
  - As the number of customers "lost" in Year 1 is simply the number of initial customers minus the number of customers who are still active at the end of the first year, we enter `=1000−D6` in Cell `E6`.
  - For $t > 1$, the number of customers "lost" in Year $t$ is the number of customers who are still active at the end of Year $t - 1$ minus the number of customers who are still active at the end of the Year $t$. We therefore enter `=D6−D7` in Cell `E7` and copy it to `E8:E12`.
- The first seven elements of the log-likelihood function are computed in Cells `F6:F12`: We enter `=E6*LN(B6)` in Cell `F6` and copy it to `F7:F12`.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | alpha | 1.000 | | | | |
| 2 | beta | 1.000 | | | | |
| 3 | LL | -2115.5 | | | | |
| 4 | | | | | | |
| 5 | t | P(T=t) | S(t) | # active | # lost | |
| 6 | 1 | 0.500 | 0.500 | 869 | 131 | -90.8 |
| 7 | 2 | 0.167 | 0.333 | 743 | 126 | -225.8 |
| 8 | 3 | 0.083 | 0.250 | 653 | 90 | -223.6 |
| 9 | 4 | 0.050 | 0.200 | 593 | 60 | -179.7 |
| 10 | 5 | 0.033 | 0.167 | 551 | 42 | -142.9 |
| 11 | 6 | 0.024 | 0.143 | 517 | 34 | -127.1 |
| 12 | 7 | 0.018 | 0.125 | 491 | 26 | -104.7 |
| 13 | | | | | | -1021.0 |

**FIGURE B1**
Screenshot of Excel Worksheet for Parameter Estimation

[4] Note that (B1) and (B2) look almost identical, but there is a subtle difference: in (B1): The probability we compute is a function of the data pattern for fixed model parameters; in (B2), we already have the data, and the probability we compute is a function of the model parameters.

- The final element of the log-likelihood function, that associated with those customers who are still active at the end of Year 7, is entered as `=D12*LN(C12)` in Cell `F13`.
- The sum of Cells `F6:F13` is entered in Cell `B3`; this is the value of the log-likelihood function given the values for the two model parameters in Cells `B1:B2`. (With starting values of 1.0 for both parameters, $LL = -2{,}115.5$.)

We find the maximum likelihood estimates of the two model parameters by maximizing the log-likelihood function. We do this using the Excel add-in Solver, available under the "Tools" menu. The *target cell* is the value of the log-likelihood, Cell `B3`. We wish to *maximize* this by *changing* Cells `B1:B2`. The *constraints* we place on the parameters are that $\alpha$ and $\beta$ are greater than 0. As Solver offers us only a "greater than or equal to" constraint, we *add* the constraint that Cells `B1:B2` are $\geq$ a small positive number (e.g., 0.0001) (see Figure B2).

Clicking the *Solve* button, Solver converges to a solution where the maximum value of the log-likelihood function is $-1{,}611.2$, associated with $\alpha = 0.668$ and $\beta = 3.806$. These

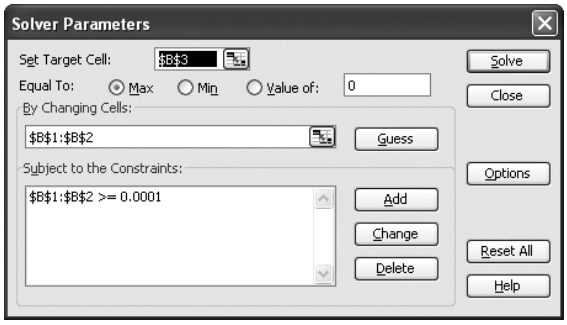


**FIGURE B2**
Solver Settings

are the maximum likelihood estimates of the model parameters. (To be sure that we actually have reached the maximum of the log-likelihood function, it is good practice to redo the optimization process using a completely different set of starting values. For example, using starting values of 0.01 and 0.01 (for which $LL = -2{,}741.7$), use Solver to find the maximum of the log-likelihood function. Are the corresponding values of the two model parameters equal to those given earlier? They should be!)